Euphytica (2005) 145: 133–146
DOI: 10.1007/s10681-005-0626-3
© Springer 2005

Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars

H. Fufa[1,4,*], P.S. Baenziger[1], B.S. Beecher[1,5], I. Dweikat[1], R.A. Graybosch[2] & K.M. Eskridge[3]
[1]*Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA;* [2]*USDA-ARS and Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA;* [3]*Department of Statistics, University of Nebraska, Lincoln, NE 68583, USA;* [4]*Present address: Monsanto Crop Genetics Research Station, 2135 W. Lincoln, Olivia, MN 56277, USA;* [5]*Present address: USDA-ARS, E-202 Food Quality Bldg., Pullman, WA 99164, USA (*author for correspondence: e-mail: fufa.h.birru@monsanto.com)*

Received 29 November 2004; accepted 7 July 2005

Key words: bread wheat, genetic diversity, *Triticum aestivum*

Summary

Genetic diversity is the basis for successful crop improvement and can be estimated by different methods. The objectives of this study were to estimate the genetic diversity of 30 ancestral to modern hard red winter wheat (*Triticum aestivum* L.) cultivars adapted to the Northern Great Plains using pedigree information, morphological traits (agronomic measurements from six environments), end-use quality traits (micro-quality assays on 50 g grain or milled flour samples for the six environments), and molecular markers (seed storage proteins separated using SDS-PAGE, 51 SSRs, and 23 SRAP DNA markers), and to determine the relationships of genetic distance estimates obtained from these methods. Relationships among diversity estimates were determined using simple (Pearson) and rank (Spearman) correlation coefficients between distance estimates and by clustering cultivars using genetic-distances for different traits. All methods found a wide range in genetic diversity. The genetic distance estimates based on pedigree had the highest values due to possible over-estimation arising from model assumptions. The genetic diversity estimates based on seed storage protein were lowest because they were the major determinants of end-use quality, which is a highly selected trait. In general, the diversity estimates from each of the methods were positively correlated at a low level with the exceptions of SRAP diversity estimates being independent of morphologic traits (simple correlation), SDS-PAGE, and SSR diversity estimates (rank correlation). However, SSR markers, thought to be among the most efficient markers for estimating genetic diversity, were most highly correlated with seed storage proteins. The procedures used to accurately estimate genetic diversity will depend largely upon the tools available to the researcher and their application to the breeding scheme.

Abbreviations: COD, coefficient of diversity; COP, coefficient of parentage; GS, genetic similarity; GD, genetic distance; GEI, genotype by environment interaction; HMW, high molecular weight; LMW, low molecular weight; SDS-PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; SSRs, simple sequence repeats; SRAPs, sequence-related amplified polymorphism

Introduction

Genetic diversity has played a vital role in the success of crop improvement. Knowledge of genetic diversity has been successfully used for efficient germplasm management and utilization, genetic fingerprinting and genotype selection (FAO, 1998; Engles et al., 2002). However, there are many methods for estimating genetic diversity. Morphological traits (syn. phenotypic traits) are commonly used to analyze genetic diversity since they provide a simple way of quantifying genetic variation while assessing genotype performance under normal growing environments. However, morphological traits are limited in number, modified by the environment and may be controlled by epistatic and pleiotropic gene effects (van Beuningen & Busch, 1997b). Despite these limitations, morphological traits have been successfully used for genetic diversity analyses and cultivar development.

In addition to agronomic performance, cultivar development with acceptable end-use quality is a major objective of winter wheat (*Triticum aestivum* L.) breeding programs (Baenziger et al., 2001). Though end-use quality is a complex characteristic, several traits have been identified for predicting relative end-use quality for milling and bread making. These traits include SDS-sedimentation value, grain protein composition, and Mixograph time and tolerance (Baenziger et al., 2001). Though environment influences most end-use quality traits, these traits are genetically controlled and can be used in selection. Similar to morphological traits, end-use quality traits may be useful for genetic diversity studies and predicting genetic variation in a breeding program.

In addition to studying diversity with morphological traits, pedigree information is used to compute pairwise coefficients of parentage (COP), which have been used as an inexpensive indicator of genetic diversity for cultivars of self-pollinating species with known pedigrees (Almanza-Pinzon et al., 2003; Murphy et al., 1986; Souza & Sorrells, 1989). The COP is an indirect estimate of the genetic relationship between two cultivars based on the probability that a random allele taken from a random locus in cultivar X is identical by descent to a random allele taken from the same locus in cultivar Y (Cox et al., 1985). Coefficient of parentage estimates are limited by errors describing the pedigree and the assumption that ancestral lines are unrelated. Furthermore, COP analyses may over- or under-estimate genetic similarity between cultivars due to selection and re-selection biases (Almanza-Pinzon et al., 2003; Cox et al., 1986; Souza & Sorrells, 1989). However, COP analyses can indicate cultivars less likely to possess similar genes (May et al., 1995). In wheat, extensive pedigree information is available and can be used to estimate genetic relationships and diversity in winter wheat cultivars and parental germplasm (Murphy et al., 1986; van Beuningen & Busch, 1997a) and to identify parents that have contributed to yield improvement (Beer et al., 1997).

A fourth way to estimate genetic diversity is with molecular markers. Molecular marker analyses overcome many of the limitations of morphological and end-use quality traits, and pedigree information-based genetic diversity analysis (Gupta et al., 1999). The most commonly used molecular markers in wheat historically have been seed storage proteins (glutenins and gliadins). They have been recommended as reliable genetic markers to differentiate wheat genotypes for breadmaking (Payne & Lawrence, 1983; Graybosch, 1992). HMW glutenin subunits are encoded by *Glu-A1*, *Glu-B1* and *Glu-D1* on the long arms of chromosomes 1A, 1B and 1D (Payne, 1987) and the LMW glutenin subunits are encoded by *Glu-A3*, *Glu-B3* and *Glu-D3* on the short arms of the same chromosomes (Gupta & Shepherd, 1990). The large gliadin families, mainly gamma and omega gliadins, are encoded by the *Gli-1* loci on the short arms of group 1 chromosomes, tightly linked to the LMW-glutenin genes at the *Glu-3* loci (Tatham & Shewry, 1995) and by the *Gli-2* loci found on the short arms of group 6 chromosomes. Genetic diversity among cultivars based on seed storage proteins is measured only in relation to chromosome groups 1 and 6 and the variation may be limited because end-use quality is a highly selected trait in many programs.

Simple sequence repeats (SSRs) are common, informative molecular markers used for genetic diversity studies because of their simplicity, high levels of polymorphism (Plaschke et al., 1995; Huang et al., 2002), high reproducibility, and co-dominant inheritance patterns (Roder et al., 1998). These markers are chromosome-specific (often amplifying a single locus with multiple alleles), can be evenly distributed along different chromosomes (Roder et al., 1998), and can be used by researchers to tag useful genes. Numerous wheat SSR markers are available and many have been mapped to specific chromosome arms (Bryan et al., 1997; Roder et al., 1998; U.S. Wheat and Barley Scab Initiative, 2003). Consequently, SSR markers are excellent markers for genetic diversity analyses and genotype identification in self-pollinated species such as wheat (Domini et al., 2000). SSR markers have been used to estimate genetic diversity in wheat germplasm (Manifesto et al., 2001; Huang et al., 2002), elite lines (Kim & Ward, 1997; Prasad et al., 2000) and cultivars (Plaschke et al., 1995; Bohn et al., 1999).

In addition to SSR markers, sequence-related amplified polymorphism (SRAP) is a new molecular marker system for genetic diversity studies in plants. SRAP is a PCR-based DNA marker system that generates multiple fragments in a single PCR reaction (Li & Quiros, 2001). SRAPs amplify several reproducible and polymorphic loci and alleles, and they may amplify functional genes since they are sequence related. As opposed to SSR markers, which tag single multiallelic loci, SRAP markers possess multiloci and multiallelic features, which make them potentially more efficient for genetic diversity analysis, gene mapping and fingerprinting genotypes. However, SRAP markers may not be randomly distributed across the genome (Li & Quiros, 2001). Limited information is available on the chromosomal locations of SRAP markers, their linkage with plant traits, and the potential of SRAP markers for genetic diversity studies in wheat. Therefore, SRAP markers were employed to examine their potential for genetic diversity analyses in hard red winter wheat.

Though there are many methods to measure genetic diversity, few studies have examined the relationships between them. Comparing results from different genetic diversity methods could identify the best method for parental selection for plant breeders (Barrett et al., 1998), thus increasing breeding efficiency. The relationships between diversity estimates can be assessed using scatter plots, correlation, regression and principal coordinate plots (Weir, 1996). The objectives of this study were to determine the genetic diversity in ancestral lines of modern hard red winter wheat cultivars adapted to the Northern Great Plains using pedigree information, morphological traits, end-use quality traits, and molecular markers, and to determine the relationships of genetic distance estimates obtained from these methods. The chosen methods of determining genetic diversity were pedigree, morphological, and end-use quality traits as these are readily available to most breeding programs. Among the molecular markers, seed storage protein composition is often available. Genetic diversity data for SSR and SRAP markers are generally not available unless the goal is to understand diversity. In addition, end-use quality and related seed storage proteins are highly selected traits, whereas SSRs markers are generally considered neutral traits. Hence, the comparisons of genetic diversity estimates in this study involve estimates from readily available data to estimates that require additional work for their derivation, and estimates from highly selected traits to estimates from neutral traits.

Materials and methods

Plant materials

The study consisted of 30 hard red winter wheat cultivars representing historically the most important ancestral lines and commonly grown cultivars over the last 130 years in the Northern Great Plains (described in detail in Fufa et al., 2005). ‘Turkey’, ‘Red Chief’, ‘Kharkof’, ‘Cheyenne’ and ‘Wichita’ are historically important parents for many subsequent hard red winter wheat cultivars. ‘Scout 66’, ‘Baca’, and ‘Eagle’ are direct pure line selections from ‘Scout’. ‘Sturdy’, ‘TAM 107’, ‘TAM 200’, ‘Colt’, ‘Karl 92’ and ‘Chisholm’ are important semidwarf cultivars. ‘Arapahoe’ (Baenziger et al., 1989) is a major parent for recent cultivars such as ‘Millennium’ and ‘Wahoo’ and was the most widely grown cultivar in Nebraska in the early 1990s. ‘Pronghorn’ is a recent high yielding and tall cultivar (Baenziger et al., 1999).

Pedigree history was traced for each cultivar to compute the pairwise COP between cultivars using the KIN software program (Tinker & Mather, 1993) to determine the degree of genetic similarity among cultivars. The coefficient of diversity (COD) calculated as 1-COP was used as a measure of the dissimilarity of parentage among cultivars. The COP computation assumed that each cultivar is completely inbred (homozygous) and homogeneous. Cultivars without a known common ancestor are assumed to be unrelated (COP = 0) and parents are assumed to contribute alleles equally to the offspring despite selection (Cox et al., 1985). Furthermore, the COP calculation considers that cultivars derived by backcrossing at least five times to their recurrent parents were genetically equivalent to their parents (Kempthorne, 1969).

Field evaluation, morphological and end-use quality characterization

The 30 cultivars were grown at Lincoln, Mead and North Platte, Nebraska during the 2002 and 2003 seasons (described in detail in Fufa et al., 2005). The morphological traits used in this study were days-to-anthesis, plant height, grain yield and grain yield components, and grain volume weight. The traits used for determining end-use quality were: flour yield, flour protein content, Mixograph mixing time and tolerance (hereafter referred to as mixing time and tolerance), and SDS sedimentation volume (Fufa et al., 2005).

Molecular marker analyses

SDS-PAGE was performed using the procedure of Graybosch and Morris (1990). Glutenins were separated in a vertical and discontinuous SDS-PAGE using a Tris–HCl buffer system. Gliadin subunits were separated on 11% acrylamide gel. The bands of HMW-

GS on SDS-PAGE were scored using the standardized methodology and nomenclature described by Payne and Lawrence (1983). The identities of HMW glutenin subunits were determined by comparison with those of standard cultivars, such as Siouxland, Abilene, Cimarron and TAM 107. Clear and reproducible gliadin subunit compositions were scored for each cultivar. Gliadins migrating between molecular weights of 21.5 and 45 kDa were scored using Scout 66 as a standard. Scout 66 gliadins in this region were scored in order of decreasing molecular weight and numbered from 1 to 11. Gliadin proteins not co-migrating with those of Scout 66 were designated by decimal or alphabetical postscripts. Wheat lines with 1BL.1RS or 1AL.1RS wheat-rye chromosomal translocations also produce secalin proteins as part of the gliadin fraction. Secalins were designated as S_1, S_2, and S_3 after Graybosch et al. (1999).

For SSRs and SRAPs, DNA extractions were carried out following the CTAB method (Saghai-Maroof et al., 1984) with slight modifications (Kuleung et al., 2004). Genomic DNA was extracted separately from a randomly selected sample of eight plants representing each cultivar and bulked at equal quantities for polymerase chain reaction (PCR) reactions. Of the 154 SSR primer pairs (Roder et al., 1998; U.S. Wheat and Barley Scab Initiative, 2003) screened for amplification and polymorphism, 68 wheat SSRs produced amplification products. The designation, chromosome-arm location and annealing temperature of SSRs used in this study was obtained from the GrainGenes database (http://wheat.pw.usda.gov, January, 2004) or Roder et al. (1998). At least one SSR marker per chromosome arm except 2AL, 4DS, 5DS, 5DL, 6AS, and 6DS was scored and a total of 51 scorable and polymorphic markers were used in the genetic diversity analysis.

SRAP marker assays followed the same procedures described for SSR markers (Kuleung et al., 2004; Roder et al., 1998; USWABSI, 2003) except that the annealing temperature for SRAP markers was changed to 48 °C. Twenty-three SRAP markers (Li & Quiros, 2001) were used for diversity estimates.

Statistical analysis

Principal component analysis was computed using PROC PRINCOMP (SAS, 1996) on correlation matrices. Each phenotypic and end-use quality attribute was standardized as follows:

$$Y_{im} = (X_{im} - \bar{X}_{im})/SD_m,$$

where X_{im} is the value for the mth morphological or end-use character before standardization ($m = 1, \ldots, M$, with $M = 10$), Y_{im} is the value for standardized character, $\bar{X}_{im}$ is the grand mean value over all cultivars of each trait, and SD_m is the standard deviation across cultivars. The individual character for each cultivar i ($i = 1, \ldots, 30$) on the variable X_{im} was an average value from six environments. The average normalized Euclidean distance between cultivar i and j (D_{ij}) using morphological or end-use quality was calculated following Roldan-Ruiz et al. (2001) as

$$D_{ij} = \sqrt{\frac{\left[\sum (Y_{im} - Y_{jm})^2\right]}{4M}}$$

where Y_{im} and Y_{jm} represent the standardized mean values for the same traits for cultivar i and j, respectively, and $4M$ ($M = 10$) is the normalizing constant ensuring that most values will be between 0 and 1. Hierarchical cluster analysis was performed using SAS PROC CLUSTER based on normalized Euclidean distance matrices with the unweighted paired group method with arithmetic averages (UPGMA) and dendrograms were constructed by PROC TREE (SAS, 1996).

The SSR and SRAP banding patterns were scored as present (1) or absent (0) for each primer pair and cultivar combination. Scored SSR products included monomorphic markers, but only polymorphic bands were considered in the genetic analysis. Gene diversity (D_L) for each locus was calculated as $D_L = 1 - \sum P_{li}^2$ and average genetic diversity (D) as a measure of genetic variation was estimated using the formula:

$$D = 1 - (1/L) \sum_l \sum_i P_{li}^2,$$

where P_i is the frequency of the ith allele at the l locus, where L is the number of loci (Weir, 1996). The genetic similarity coefficients (GS) or the Dice coefficients (Sneath & Sokal, 1973) were computed between pairs of 30 cultivars to obtain a genetic similarity matrix based on SSR and SRAP banding patterns. Genetic similarity between two cultivars within one locus was calculated using the formula

$$\mathrm{GS}_{ij} = 2N_{ij}/(N_i + N_j),$$

where N_i and N_j represent the total number of bands present in cultivars i and j, respectively, and N_{ij} refers to the total number of bands common to the same cultivars (Nei & Li, 1979). Thus, GS_{ij} reflects the proportion of bands in common between two parents and may range from 0 (no common bands) to 1 (identical profiles for two bands). Genetic similarity between two cultivars is the average similarity over all loci and the coefficient of dissimilarity between them is therefore computed as $GD_{SSR} = 1 - GS_{ij}$ (Nei & Li, 1979). The value of genetic dissimilarity coefficients is also considered to be the distance between pairs of cultivars and usually ranges from 0 (when bands in two lines are identical) to 1 (when there are no bands in common between two cultivars). Similar to SSRs, SRAP amplified fragments were scored as present (1) or absent (0) for diversity analysis. Protein subunits were also converted to a binary matrix consisting of present (1) or absent (0) for diversity analysis. Genetic diversity of a genotype is the average diversity of all loci (Nei & Li, 1979). Genetic distance matrices were used to cluster cultivars using hierarchal cluster analysis based on UPGMA and the results were used to construct dendrograms. Cluster analyses were analyzed using the same procedure based on glutenins and gliadins. All analyses of genetic diversity based on SSRs, SRAPs and protein compositions were performed by the NTSYSpc version 2.1 computer software program (Rohlf, 2000). Correlation (simple and rank) and cluster analyses based on genetic distance were used to examine the relationships among genetic diversity estimates.

Results and discussion

Genetic diversity and relationships among cultivars

The genetic distances for each method (Table 1) and the clusters based on genetic differences (Table 2) indicated that the methods varied greatly in their estimates of genetic diversity and to a lesser extent in the clusters they generated. As some methods appeared to overestimate and others underestimate genetic distances between cultivars, the threshold level at which the same number of clusters formed from different methods also varied. The threshold increased as the method overestimated the distance between cultivars. Hence, rather than using an arbitrary threshold level and arbitrary number of clusters, only meaningful clusters were considered. A very similar number of clusters (4 for morphology, seed storage proteins, and all markers; 5 for end-use quality and SRAPs, 6 for SSRs, and 7 for pedigree) with the threshold from 0.43 to 0.95 were found from different methods in this study (Table 2). Representative dendograms for the lines based on pedigree information (Figure 1), morphological information (Figure 2), seed storage proteins (Figure 3), SSRs markers (Figure 4) and combining all molecular markers (seed storage proteins, SSRs, and SRAPs; Figure 5) are presented to help visualize the clusters.

The pedigree-based genetic distance estimates between most cultivars were higher on average than other diversity measures (Table 1). Larger average distance values in pedigree-based genetic distance estimates are expected because the method does not allow intermediate values for many pairwise comparisons between zero and one, assumes that ancestors are unrelated, and assumes no selection and equal contribution of parents to the progeny. In studying and comparing the different methods of estimating genetic diversity or relatedness and the clusters, the differences among the methods can be illustrated by the five ancestral cultivars (Turkey, Cheyenne, Red Chief, Kharkof, and Wichita). Using pedigree analysis, each was placed in a separate cluster. Turkey was clustered with Scout 66 and other Scout selections and derivatives (Table 2). Cheyenne, which is believed to be the foundation of the Nebraska breeding program, was clustered with many of the modern Nebraska cultivars. Wichita was clustered mainly with lines developed in the southern Great Plains. The

Table 1. Means, standard deviation and ranges of genetic distance matrices based on pedigree, morphological traits, end-use quality attributes, proteins, SSRs, and SRAPs among 30 hard red winter wheat cultivars

Parameters	GD_{ped}[a]	GD_{moph}[b]	$GD_{quality}$[b]	$GD_{proteins}$[c]	GD_{SSRs}[c]	GD_{SRAP}[c]	$GD_{allmolmarkers}$[c]
Mean	0.895	0.500	0.499	0.245	0.427	0.357	0.359
Maximum	1.000	1.917	1.997	0.586	0.698	0.677	0.524
Minimum	0.125	0.060	0.018	0.000	0.171	0.111	0.141
Standard deviation	0.152	0.311	0.397	0.118	0.082	0.094	0.067

[a]GD_{ped} is genetic distance based on pedigree, calculated as 1-KIN coefficient (Tinker & Mather, 1993).[b] GD_{morph}, and GD_{qual}, are genetic distances obtained from morphology and end-use respectively and calculated as described by Roldan-Ruiz et al. (2001).[c]$GD_{proteins}$, GD_{SSRs}, GD_{SRAP} and $GD_{allmolmarkers}$ are genetic distances obtained from proteins, SSRs, SRAPs and all molecular markers (proteins, SSRs, and SRAPs), respectively, calculated based on Nei and Li (1979).

Table 2. List of hard red winter wheat cultivars in the major clusters based on pedigree, morphology, end-use quality trait and molecular markers. Ancestral cultivars (Turkey, Cheyenne, Red Chief, Kharkof, and Wichita) are in bold

Trait	Cluster	Cultivars
Pedigree	I	**Turkey**, Scout 66, Eagle, Baca, Sage, Buckskin, Bennett and Nekota
	II	Colt, Pronghorn and TAM 200
	III	**Cheyenne**, Warrior, Siouxland, Redland, Centurk 78, Centura, Arapahoe, Wahoo Culver and Millennium
	IV	**Wichita**, Sturdy, Chisholm, Alliance, TAM 107 and Niobrara
	V	Karl 92
	VI	**Red Chief**
	VII	**Kharkof**
Morphology	I	**Turkey, Kharkof, Red Chief, Wichita** and **Cheyenne**
	II	Warrior, Scout 66, Buckskin, Baca, Eagle, Sage, Bennett, Arapahoe, Centura, Pronghorn, Sturdy, Niobrara, Wahoo, Culver, Millennium, Centurk 78, Siouxland, Colt, Nekota, Chisholm, Karl 92, TAM 107 and Alliance
	III	TAM 200
	IV	Redland
End-use quality	I	**Turkey, Wichita, Kharkof** and **Red Chief**
	II	Sturdy, Siouxland, Millennium and Alliance
	III	TAM 200
	IV	**Cheyenne**, Baca, Bennett, Warrior, Scout 66, Colt, Sage, Nekota, TAM 107, Culver, Redland, Arapahoe and Niobrara
	V	Eagle, Karl 92, Buckskin, Centurk78, Centura, Pronghorn and Chisholm
Proteins	I	**Turkey**, Centurk 78, Centura, Karl 92, **Kharkof**, Warrior, Eagle, Sage, Redland, Arapahoe and **Wichita**
	II	**Cheyenne**, Pronghorn, Bennett, Alliance, Culver, Buckskin, Millennium, Wahoo, Colt, Chisholm, Sturdy, and Scout 66
	III	Siouxland, TAM 107, Nekota, Niobrara, and TAM 200
	IV	**Red Chief**
SSR	I	**Turkey, Wichita**, Centura, Pronghorn, Centurk78, Millennium and Buckskin
	II	Scout 66, Baca, Eagle, Sage, Bennett, Nekota, Culver, Redland, Niobrara, Alliance, Colt, TAM 107, TAM 200, Chisholm, Arapahoe and Wahoo
	III	**Kharkof, Cheyenne** and **Red chief**
	IV	Warrior and Siouxland
	V	Sturdy
	VI	Karl 92
SRAP	I	**Turkey,** Scout 66, Bennett, Nekota, Redland, Arapahoe, Millennium and Niobrara
	II	Baca, Sage, Alliance, Buckskin, Colt, Chisholm, Karl 92, Siouxland, Culver, Wahoo and TAM 107
	III	**Red Chief, Cheyenne, Wichita**, Sturdy and TAM 200
	IV	Centurk78, Centura and Pronghorn
	V	**Kharkof**, Eagle and Warrior
All markers[a]	I	**Turkey**, Centurk78, Millennium, **Wichita**, Pronghorn, Centura, Siouxland, Buckskin and Warrior
	II	**Kharkof**, **Cheyenne, Red Chief**, Sturdy and Karl 92
	III	Scout 66, Baca, Eagle, Sage, Bennett, Nekota, Culver, Redland, Niobrara and Alliance
	IV	Colt, Chisholm, Arapahoe, Wahoo, TAM 107 and TAM 200

[a] All marker uses the marker information from SSRs, SRAPs, and proteins.

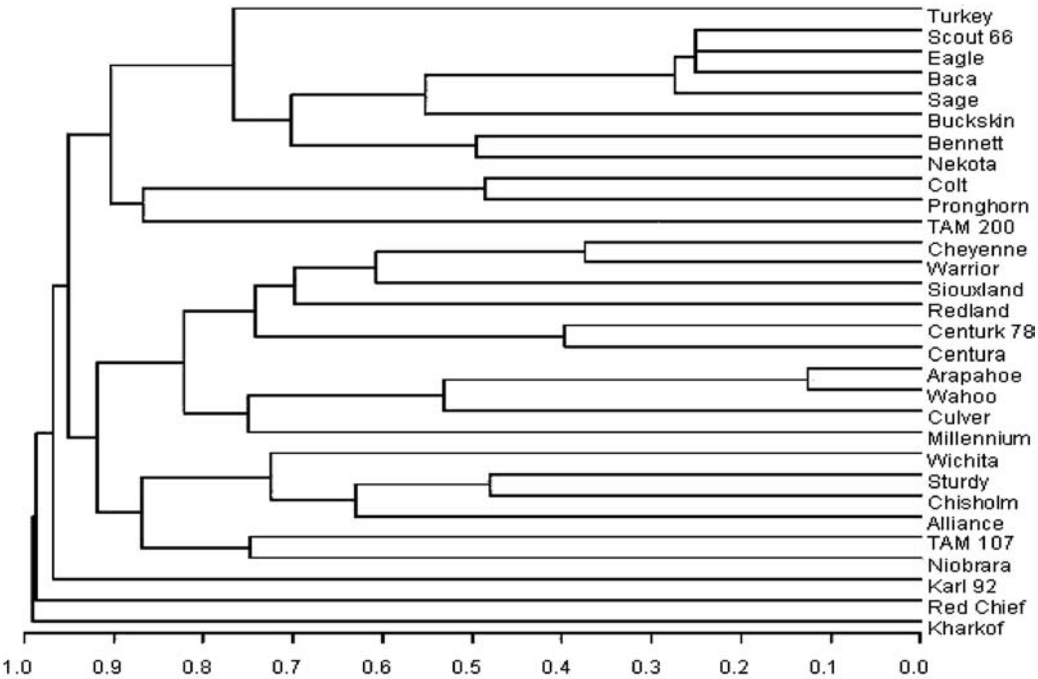

Figure 1. Dendrogram of 30 hard red winter wheat cultivars as revealed by pedigree information using the average method of clustering.

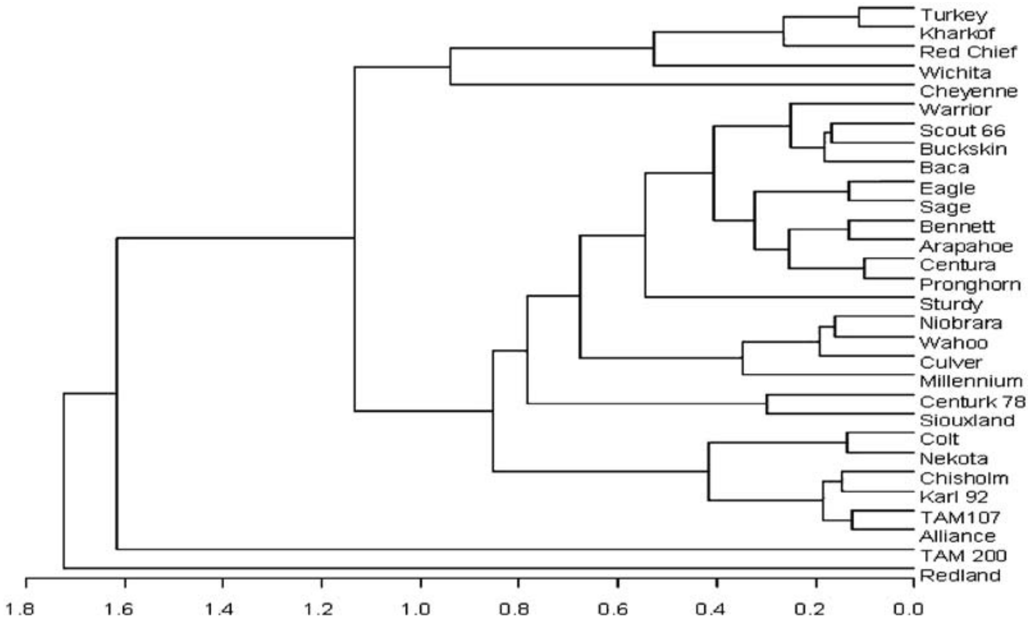

Figure 2. Dendrogram of 30 hard red winter wheat cultivars based on 10 morphological characters measured at six Nebraska environments using the average method of clustering.

remaining clusters were small. The distance estimates between ancestral cultivars and most unrelated cultivars were the highest ($GD_{ped} = 1.00$) and assuming no selection (which would cause similarly selected cultivars to be more related) would have inflated genetic diversity values as was previously reported (Cox et al., 1985; Murphy et al., 1986; van Beuningen & Busch, 1997a; Almanza-Pinzon et al., 2003). These limitations

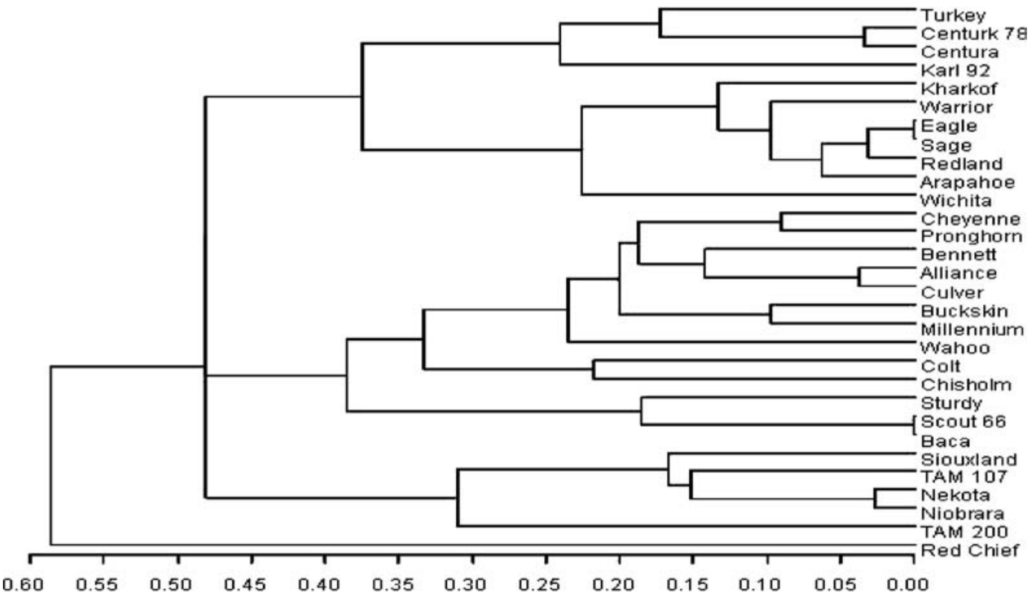

Figure 3. Dendrogram of 30 hard red winter wheat based on genetic distance from proteins (secalins, glutenins and gliadins) using average method of clustering.

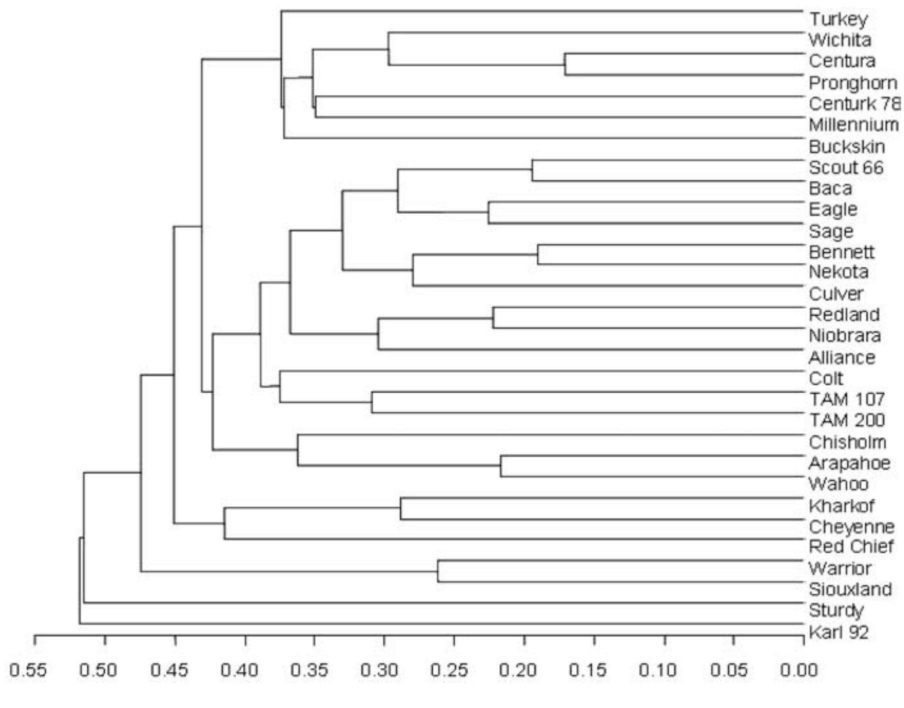

Figure 4. Dendrogram of 30 hard red winter wheat cultivars using SSR markers as per the average method of clustering.

force breeders to consider other methods for estimating genetic diversity in wheat.

Morphological traits, involving principal component analyses, are also commonly used in genetic diversity estimates. Principal component analyses of morphological traits (data not shown) found that the first principal component, which explained 34% of the total variability among cultivars, contrasted plant

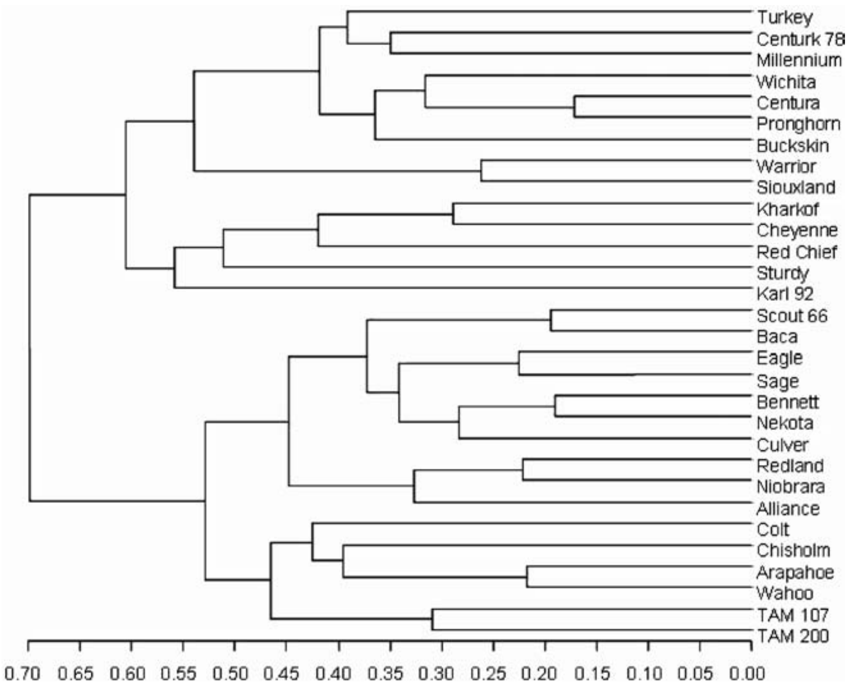

Figure 5. Dendrogram of 30 hard red winter wheat cultivars based on SSR, SRAPs and proteins (glutenins, gliadins and secalins) as per average method of clustering.

height and culm length with the grain yield and the number of spikes per square meter. The result implies that cultivars characterized by short height were higher yielding as expected because there are few modern tall cultivars and no historic semidwarf cultivars. The second principal component explained 25% of the total variability and indicated the joint importance of grain weight per spike and number of kernels per spike in discriminating cultivars. The morphological trait-based distance values used to estimate the genetic relationships between pairs of cultivars ranged from 0.06 (Pronghorn and 'Centura') to 1.92 between 'Redland' and TAM 200 with an average distance of 0.50 (Table 1). The morphology clusters were relatively uninformative (effectively two clusters – the ancestral lines, and the non-ancestral cultivars, plus two single cultivar clusters); however the dendogram was more informative by having additional subclusters (Figure 2).

Grain end-use quality is an important trait for the commercial value of a cultivar. Microquality-based variability analyses (data not shown) depicted that the first PC identified SDS-sedimentation value, mixing time and tolerance as the most important traits accounting for 46% of the differences among cultivars whereas the second component identified protein content and accounted for 24% of the differences among cultivars. Distance estimates based on five microquality traits ranged from 0.018 to 1.997 with an average value of 0.499 (Table 1), which was very similar to that of the morphological traits. The end-use quality clusters also highlighted the ancestral cultivar similarities as four of the five ancestral cultivars were in one cluster mainly due to their lower end-use quality by current standards, particularly low mixing time and tolerance and SDS sedimentation. The two larger clusters were: (1) Cheyenne and Scout with their derivatives having intermediate mixing times and tolerances and (2) cultivars with long mixing times and tolerances (e.g. Karl 92, Pronghorn and Centura).

Seed storage proteins (allelic compositions for genes encoding high molecular weight glutenins and gliadins) were determined for the cultivars. Similar to results of Graybosch (1992), allelic variation was identified among cultivars at three high molecular weight glutenin loci. At *Glu-A*, the allele encoding subunit 2* was found in 26 cultivars, whereas subunit 1 was in 2 cultivars, and 2 cultivars were heterogeneous. At *Glu-B1*, four allelic combinations (subunits 6 + 8, 7 + 8, 7 + 9, 13 + 16) were identified. The 7 + 9 subunits were found in 23 cultivars and were heterogeneous in two. The 7 + 8 subunits were identified in four cultivars and one heterogeneous cultivar. The rare subunits 6 + 8 and 13 + 16 (Graybosch, 1992) were identified only in Red Chief (6 + 8) and Wahoo (heterogeneous for 13 + 16 and 7 + 9). At *Glu-D1*, the 5 + 10 subunits associated with good breadmaking quality (Graybosch, 1992; Payne & Lawrence, 1983) were found in all cultivars except Sturdy, Scout 66 and Baca, with subunits 2 + 12, which are generally associated with poorer bread-making quality. Cultivars varied in their gliadin subunit composition, as they did for glutenins. With reference to Scout 66, cultivars were scored for 18 reproducible and polymorphic gliadin proteins. Secalins were derived from genes located on 1RS rye chromosome segments and were identified only in Niobrara (1AL.1RS), Nekota (1AL.1RS), TAM 200 (1AL.1RS), TAM 107 (1AL.1RS), and Siouxland (1BL.1RS), which clustered together.

The average genetic distance estimate (0.245), with the range of 0.00–0.586 among 30 winter wheat cultivars revealed by seed storage proteins was the lowest of all diversity estimates (Table 1). The genetic distance estimates between a few cultivars, viz. Baca and Scout 66, and Sage and Eagle, were zero due to their possession of identical high molecular weight glutenin and gliadin subunit combinations. This result indicates the power of selection (Baenziger et al., 2001) to lessen the genetic diversity of seed storage proteins, especially for a trait that is controlled by a few major genes. The seed storage protein clusters also highlighted the similarities of ancestral cultivars as three of the five cultivars were in one cluster. Turkey, a heterogeneous line for storage proteins (Graybosch, 1992), was clustered with Centurk 78 and Centura in which glutenin subunits 7 + 8 were identified. The remaining cultivars in the same cluster were related as they possess at least 2*, 7 + 9 and 5 + 10, no secalins and similar gliadin banding patterns. The second cluster (45% of cultivars) consisted of Cheyenne, Scout selections and most modern Nebraska cultivars due to their possession of protein subunits associated with high breadmaking quality (2*, 7 + 9 and 5 + 10). Red Chief clustered by itself (Table 2, Figure 3).

In addition to seed storage proteins, several other molecular markers have been developed for diversity and related studies (Prasad et al., 2000). Sixty-eight of 154 wheat SSR markers screened for amplification products and polymorphism information produced a total of 141 bands (both monomorphic and polymorphic bands) across all cultivars (data not shown). The SSR markers were more polymorphisms than reported in previous studies (Plaschke et al., 1995; Bohn et al., 1999). The average number of bands per locus was 3 bands (ranged from 1 to 5). Gene diversity per locus ranged from 0.289 to 0.958 and the average genetic distance across all loci in 30 cultivars was 0.623. The average genetic distance was 0.427 (Table 1).

SSR markers also revealed higher genetic distances compared to seed storage proteins between most pairs of cultivars indicating that the cultivars were developed from diverse germplasm. A similar result was obtained by Cox et al. (1986), who reported that genetic diversity has increased in hard red winter wheat as opposed to other wheat classes. In 22 very diverse wheat genotypes, Sun et al. (1998) observed low RAPD-based genetic distance estimates ranging from 0.062 to 0.340 with an average of 0.179. The higher SSR-based distance observed in our study could be due to more complete coverage of the genomes with markers or to the diversity of the lines used in the study. Using a more diverse set of cultivars, Almanza-Pinzon et al. (2003) found higher levels of diversity (their data are reported as similarity). In this study, the ancestral cultivars were in two clusters. One cluster (I) consisted of Turkey and Wichita plus mainly other traditional tall wheat cultivars. Kharkof, Cheyenne, and Red Chief were in another cluster (III). Most of the remaining cultivars, Scout 66 or Brule selections and their derivatives, were in Cluster II.

Twenty-three SRAP markers produced 468 amplified fragments (including 60 monomorphic fragments) with an average genetic diversity of 0.418 and range of 0.10–0.90. The SRAP marker assay revealed up to three reproducible loci and 7 alleles per marker (data not shown). The diversity estimates for SRAP markers ranged from 0.11 to 0.677 with an average value of 0.357. As such, the SRAP markers provided more conservative estimates of genetic diversity than SSR markers. SRAP marker-based clustering (Table 2) was noticeable for its differences from the SSR clusters, again indicating that it may be measuring a different aspect of genetic diversity. For example, using the SRAP clusters, Turkey (Cluster I) was not grouped with any ancestral cultivar, Buckskin (Cluster II), Wichita (Cluster III), Centura (Cluster IV), or Kharkof (Cluster V) as it was in the SSR clusters. Similarly, the Scout derivatives were placed in three different clusters, Cluster I (Scout 66), Cluster II (Baca), and Cluster V (Eagle). Red Chief, Cheyenne, and Wichita were clustered together as were some related cultivars (e.g. Bennett and Nekota, Pronghorn and Centura, and Arapahoe and Redland clustered together). These results indicate that

SRAP markers have potential for genetic diversity and genotype identification, but will give different clustering patterns than SSR clusters.

The combined analysis of genomic regions amplified by SSRs, SRAP and seed storage proteins gave genetic distance estimates that averaged 0.359 and ranged from 0.141 to 0.524 and were most similar to the estimates from SRAPs. The combined analysis may have avoided the underestimates of genetic distances based on seed storage proteins and overestimates of genetic distances based on pedigree. The fact that combined diversity estimates based on several molecular markers cover more genomic regions than a single marker alone, genetic distance estimates based on all molecular markers most likely gave the most unbiased distance estimates. Using all of the molecular markers (SSRs, SRAPs and seed storage proteins), Turkey and Wichita clustered together with mainly tall wheat cultivars. The remaining ancestral cultivars (Kharkof, Cheyenne, and Red Chief), clustered with Sturdy and Karl 92. Cluster III was mainly Scout or Brule selections and derivatives. Cluster IV included many modern semidwarf cultivars. Most clusters formed from clustering using all molecular markers were largely based upon the contributions of the SSR markers. The only unexpected placement of cultivars included the clustering of Chisholm and Colt with Arapahoe and Wahoo, which were derived from Brule (Cluster IV) and not with Alliance, which had both lines as parents, and the placement of Centurk 78 with Millennium (Cluster I). These clusters were influenced by the diversity in seed storage protein composition, a highly selected trait, rather than other markers.

In comparing the clusters and the methods of estimating genetic diversity or relatedness, the differences again can best be illustrated by the five ancestral cultivars (Turkey, Cheyenne, Red Chief, Kharkof, and Wichita). Using pedigree analysis, each was in a separate cluster. However, all five were clustered together by morphology, in two clusters (end-use quality, SSRs, and the combined molecular markers), or three clusters (seed storage proteins, SRAPs). Clearly, pedigree analysis overestimated diversity because the ancestral cultivars were assumed to be unrelated if they had no known common parent.

Correlations between genetic diversity estimates of cultivars

With the exception of distance estimates based on morphological traits and SRAPS (simple), seed storage proteins with SRAPs (rank), and SSRs with SRAPS (rank), all of the genetic distances were positively correlated. However, most correlation values were low indicating they explained little of the variation detected by the other methods of estimating genetic distances.

In general, the correlation between genetic distance estimates from pedigree and all other diversity estimates tended to be low, probably due to the unrealistic genetic assumptions made for calculating COP (Table 3). Similarly, the correlations between the genetic distance estimated by morphological traits and the other methods were positive, but also low. This positive correlation is important because morphological traits continue to be an efficient way of routinely evaluating several lines derived in a breeding program for breeders without access to laboratory-based assessment tools. The highest correlation was with end-use quality traits and may be explained by morphology and end-use quality diversity estimates both being based on highly selected traits for their estimates, or by the limited number of lines used in this study.

As expected, end-use quality diversity estimates were most highly correlated with seed storage proteins, the proteins which have a major impact on end-use quality (Payne, 1987; Graybosch, 1992). Somewhat surprisingly, the genetic distance estimates from seed storage proteins were almost equally correlated to end-use quality as with those from SSRs markers, which are assumed to be neutral under selection and were chosen to sample more of the genomes as opposed to two homoeologous chromosome groups. This result may indicate that while many seed storage proteins are retained by selection, others, especially the gliadins, were neutral in selection similar to SSR markers. The positive correlation between the genetic distance estimates from seed storage proteins and SSR markers suggested that both methods could be used to estimate diversity among cultivars, though SSR markers are preferred because selection has less impact and more of the genome is evaluated. Positive correlation between molecular markers-based genetic distances was expected as previously reported (Almanza-Pinzon et al., 2003; Bohn et al., 1999; Powell et al., 1996). Some authors reported significant positive correlation of SSR with AFLP (Almanza-Pinzon et al., 2003; Bohn et al., 1999) and SSR with COP (Almanza-Pinzon et al., 2003; Bohn et al., 1999; Kim & Ward, 1997) in wheat, though limited information is known about its correlation with morphological data.

The low, positive correlation between diversity estimates from SRAPs and all other methods perhaps

Table 3. Simple and rank correlation coefficients ($n = 435$) between distance matrices generated using morphological traits, end-use quality attributes, SSRs, SRAPs, proteins, and pedigree information of 30 hard red winter wheat cultivars

	Simple and rank correlation coefficients between distance estimates[a]					
Distance measures	End-use quality traits	Proteins (SDS-PAGE)	SSRs	SRAPs	SSRs, SRAPs and SDS-PAGE	Pedigree information
Morphologic traits	0.29**	0.22**	0.21**	0.09	0.28**	0.21**
	0.27**	0.21**	0.22**	0.11*	0.27**	0.13**
End-use quality traits		0.44**	0.25**	0.21**	0.38**	0.23**
		0.40**	0.22**	0.17**	0.34**	0.20**
Proteins (SDS-PAGE)			0.42**	0.11*	0.64**	0.28**
			0.42**	0.09	0.61**	0.24**
SSRs				0.12*	0.76**	0.28**
				0.08	0.72*	0.18**
SRAPs					0.46**	0.15**
					0.45**	0.11*
SSRs, SRAPs and SDS-PAGE						0.36**
						0.17**

[a]Upper and lower figures simple (Pearson) and rank (Spearman) correlation coefficients, respectively. *,**Significant and highly significant correlation at 5 and 1% probability level, respectively.

reflects that SRAPs are sequence related amplified markers and may measure molecular diversity differently from other diversity measures used in this study. It should be noted that the genetic diversity estimated from SRAPs was lower than all other estimates except those from the highly selected seed storage proteins.

The genetic distance estimates using all of the markers (seed storage proteins, SSRs, and SRAPs) were, as expected, highly correlated to the estimates from seed storage proteins, SSRs and SRAPs as they were contributors to the combined distance estimate. However, the combined marker genetic diversity estimates were also highly correlated to the genetic diversity estimates based on morphological, end-use quality traits, and pedigree indicating the general utility of this genetic diversity estimate.

In summary, genetic diversity analyses indicated that hard red winter wheat cultivars adapted to Northern Great Plains were diverse for morphologic, end-use quality, and molecular markers. Molecular marker-based genetic diversity estimates between cultivars revealed inflated pedigree-based genetic distance estimates between most cultivars perhaps due to the assumptions for estimating coefficient of parentage. The lowest genetic diversity was estimated from seed storage proteins, a trait which was indirectly selected via selection for end-use quality. Genetic diversity based on end-use quality was considerably higher which may reflect the difference between a trait that is affected by environment (end-use quality) and one that is not (seed storage proteins). Alternatively, it may indicate that end-use quality is affected by more of the genome than is sampled by glutenin, gliadin, or secalin bands. Genetic distance estimates from seven methods of estimating genetic diversity were generally correlated, but at a low level. There were consistencies and inconsistencies in the cultivar clusters showing their interdependence and complimentary nature. Any one of these methods could be used to study diversity and group genotypes, but none would be fully interchangeable in use. The choice of genetic diversity estimate will depend largely upon the tools available to the researcher and how they fit into the breeding scheme. For example, if SDS-PAGE is routinely done, it may be the method of choice or the method to augment genetic diversity based on pedigree information though it will underestimate genetic diversity compared to DNA marker estimates. Because the SRAP markers tended to have low correlation with the other genetic diversity estimates, they may provide different and unique insights into genetic diversity.

References

Almanza-Pinzon, M.I., M.L. Warburton, P.N. Fox & M. Khairallah, 2003. Comparison of molecular markers and coefficients of parentage for the analysis of genetic diversity among spring bread wheat accessions. Euphytica 130: 77–86.

Baenziger, P.S., D.R. Shelton, M.J. Shipman & R.A. Graybosch, 2001. Breeding for end-use quality: Reflections on the Nebraska experience. Euphytica 119: 95–100.

Barrett, B.A., K.K. Kidwell & P.N. Fox, 1998. Comparison of AFLP and pedigree-based genetic diversity assessment methods using wheat cultivars from the Pacific Northwest. Crop Sci 38: 1271–1278.

Beer, S.C., W. Siripoonwiwat, L.S.O. Donoughu, E. Souza, D. Mattews & M.E. Sorrells, 1997. Association between molecular markers and quantitative traits in oats germplasm pool: Can we infer linkage? http://www.cabi-publishing.org/jag/index.html.

Bohn, M., H.F. Utz & A.E. Melchinger, 1999. Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. Crop Sci 39: 228–237.

Bryan, G.J., P. Stephenson, A.J. Collins, J. Kirby, J.B. Smith & M.D. Gale, 1997. Low levels of DNA sequence variation among adapted genotypes of hexaploid bread wheat. Theor Appl Gent 99: 192–198.

Cox, T.S., J.P. Murphy & D.M. Rodgers, 1985. Coefficient of parentages for 400 winter wheat cultivars. In: Agricultural Experiment Station. Kansas State University, Manhattan, USA.

Cox, T.S., J.P. Murphy & D.M. Rodgers, 1986. Changes in genetic diversity in the red winter wheat regions of the United States. Proc Natl Acad Sci USA 83: 5583–5586.

Domini, P., J.R. Law, R.M.D. Koebner, J.C. Reeves & R.J. Cooke, 2000. Temporal trends in the diversity of UK wheat. Theor Appl Genet 100: 912–917

Engles, J.M.M., V.R. Rao, A.H.D. Brown & M.T. Jackson, 2002. Managing Plant Genetic Diversity, p. 487. CABI Publishing, UK.

FAO, 1998. The States of the World's Plant Genetic Resources for Food and Agriculture, p. 510. FAO, Rome, Italy.

Fufa, H., P.S. Baenziger, B.S. Beecher, R.A. Graybosch, K.M. Eskridge & L A. Nelson, 2005. Genetic improvement trends in agronomic performances and end-use quality characteristics among hard red winter wheat cultivars in Nebraska. Euphytica (in press).

Graybosch, R.A., 1992. High molecular weight glutenin subunit composition of cultivars, germplasm and parents of U. S. red winter wheat. Crop Sci 32: 1151–1155

Graybosch, R.A. & R. Morris, 1990. An improved SDS-PAGE method for the analysis of wheat endosperm storage proteins. J Cereal Sci 11: 201–212.

Graybosch, R.A., J.-H. Lee, C.J. Peterson, D.R. Porter & O.K. Chung, 1999. Genetic, agronomic and quality comparisons of two 1AL.1RS wheat-rye chromosomal translocations. Plant Breed 118: 125–130.

Gupta, R.B. & K.W. Shepherd, 1990. Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutelin. 1. Variation and genetic control of the subunits in hexaploid wheats. Theor Appl Genet 80: 65–74.

Gupta, P.K., R.K. Varshney, P.C. Sharma & B. Ramesh, 1999. Molecular markers and their applications in wheat breeding. Plant Breed 118: 369–390.

Huang, X.Q., A. Borner, M.S. Roder & M.W. Ganal, 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 105: 699–707.

Kuleung, C., P.S. Baenziger & I. Dweikat, 2004. Transferability of SSR markers among wheat, rye and triticale. Theor Appl Genet 108: 1147–1150.

Kempthorne, O., 1969. An Introduction to Genetic Statistics. Iowa State University Press, Ames, IA.

Kim, H.S. & R.W. Ward, 1997. Genetic diversity in Eastern U.S. soft winter wheat (*Triticum aestivum* L. em. Thell.) based on RFLPs and coefficient of parentage. Theor Appl Genet 94: 472–479.

Li, G. & C.F. Quiros, 2001. Sequence-related amplified polymorphism (SRAP), a new marker system based on a sample PCR reaction: Its application to mapping and gene tagging in *Brassica*. Theor Appl Genet 103: 455–461.

Manifesto, M.M., A.R. Schlatter, H.E. Hopp, E.Y. Suarez & J. Dubcovsky, 2001. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci 41: 682–690.

May, O.L., D.T. Bowman & D.S. Calhoun, 1995. Genetic diversity of U.S upland cotton cultivars released between 1980 and 1990. Crop Sci 35: 1570–1574.

Murphy, J.P., T.S. Cox & D.M. Rodgers, 1986. Cluster analysis of red winter wheat cultivars based upon coefficients of parentage. Crop Sci 26: 672–676.

Nei, M. & W.H. Li, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76: 5269–5273.

Payne, P.I., 1987. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Ann Rev Plant Physiol 38: 141–153.

Payne, P.I. & G.J. Lawrence, 1983. Catalogue of alleles for the complex loci, Glu-A1, Glu-B1, and Glu-D1, which code for high-molecular-weight subunits of glutenin in hexaploid wheat. Cereal Res Commun 11: 29–35.

Plaschke, J., M.W. Ganal & M.S. Roder, 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91: 1001–1007.

Powell, W., M. Morgante, C. Andree, M. Hanafey, J. Vogel, S. Tingery & A. Rafalski, 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2: 225–238.

Prasad, M., R.K. Varshney, J K. Roy & H.S. Balyan, 2000. The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. Theor Appl Genet 100: 584–592.

Roder, M.S., V. Korzun, K. Wendehake, J. Plaschke, M. Tixier, P. Lorey & M.W. Ganal, 1998. A microsatellite map of wheat. Genetics 149: 2007–2023.

Roldan-Ruiz, I., F.A. Van Eeuwijk, T.J. Gilliland, P. Dubreuil, C. Dillmann, J. Lallemand, M. DeLoose & C.P. Baril, 2001. A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. Theor Appl Genet 103: 1138–1150.

Rohlf, F.J., 2000. NTSYS-pc Numerical Taxonomy and Multivariate Analysis System, v. 2.202k. Exeter Software, Setauket, New York.

Saghai-Maroof, M.A., K.M. Soliman, R.A. Jorgensen & R.W. Allard, 1984. Ribosomal DNA spacer-length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81: 8014–8018.

SAS Institute Inc, 1996. SAS/STAT User's Guide, Version 6, 4th edn., Vols. 1 and 2. SAS Institute, Inc., Cary, NC.

Sneath, P.H.A. & R.R. Sokal, 1973. Numerical Taxonomy. The Principles and Practices of Classification. W.H. Freeman and Co., San Francisco.

Souza, E. & M.E. Sorrells, 1989. Pedigree analysis of North American oat cultivars released from 1951 to 1985. Crop Sci 29: 595–601.

Sun, Q., Z. Ni, Z. Liu, J. Gao & T. Haung, 1998. Genetic relationships and diversity among Tibetan wheat, common wheat and European spelt wheat revealed by RAPD markers. Euphytica 99: 205–211.

Tatham, A.S. & P.R. Shewry, 1995. The S-poor prolamins of wheat, barley and rye. J Cereal Sci 22: 1–16.

Tinker, N.A. & D.E. Mather, 1993. Software for computing kinship coefficients. Heredity 84: 238.

U.S. Wheat and Barley Scab Initiative, 2003. Preliminary BARC SSR maps and primer pair. http://www.scabusa.org/pdfs/BARC-SSR.html.

van Beuningen, L.T. & R.H. Busch. 1997a. Genetic diversity among North American spring wheat cultivars: I. Analysis of the coefficients of parentage matrix. Crop Sci 37: 570–579.

van Beuningen, L.T. & R.H. Busch, 1997b. Genetic diversity among North American spring wheat cultivars: III. Cluster analysis based on quantitative morphological traits. Crop Sci 37: 981–988.

Weir, B.S., 1996. Genetic Analysis. II. Sinauer Publishers, Sunderland, MA, USA.